Coming Together to Raise Achievement

New Assessments for the Common Core State Standards

February 2011

Prepared by the
Center for K–12 Assessment & Performance Management at ETS

Table of Contents

Center for K–12 Assessment & Performance Management at ETS

Pascal (Pat) D. Forgione Jr., Ph.D.
Distinguished Presidential Scholar and Executive Director

Dear Colleague:

How fortunate we are to be participants in a historic moment in education reform. The federal Race to the Top initiative has brought forward unprecedented resources to stimulate bold proposals for improving the way our schools operate and the quality of tools available to our students and educators. Used well, it is a much-needed investment in our nation's future. While the performance of U.S. students has improved over the past two decades, the pace has been too slow. Student performance in major competitor countries is improving at a faster rate — and some have already surpassed us. We must reverse that trend, and do so quickly.

One critical strategy in this effort is the Race to the Top Assessment Program, which has provided funding to two Consortia of states to develop a new generation of assessments intended to yield timely data to enhance instruction, accelerate learning, and provide accurate information on how our students and schools are performing. Each Consortium has more than $175 million and just four short years to push the frontiers of the assessment field and build their new testing and instructional support systems.

This initiative is about much more than just "better tests," however. The adoption of a common set of college- and career-readiness standards by 43 states and the District of Columbia, to date, means that more than 80 percent of our nation's public school students and teachers will be focused on the same content standards for their students. This is a fundamental shift in the education marketplace. As Thomas Friedman, author of *The World is Flat*, noted, innovations that allow large numbers of people to access one another's ideas and collaborate have, in the past, created "inflection points" — marked increases in the pace of innovation. The Common Core State Standards (CCSS), aligned common assessments, and the open-source platforms being developed to collect and share resources create the opportunity for an inflection point in American public education.

The Center for K–12 Assessment & Performance Management at ETS® has created this guide to stimulate discussions about the opportunity before us. We begin with an overview of the CCSS — what's new, why they're needed, and how these standards will drive new designs in assessment. We then provide you with descriptions of the two assessment systems to be built. Finally, we provide our best thinking on the work ahead in assessment — the most significant and high-leverage measurement and technical challenges within the designs of these new assessment systems.

This is truly a time that calls for a groundswell of American ingenuity and a fierce commitment to finding solutions. As the income gap between the well-educated and the poorly educated grows and we struggle to rebuild our economy, it has become clear that improving public education is both the moral and the economic imperative of our generation. Let's each do our part.

WHAT WILL THE NEW ASSESSMENTS MEASURE?

Our nation's current patchwork of state standards has resulted in great variability in the academic expectations for students. A student found to be performing at the distinguished level in one state may be below the standard in another. This doesn't make sense — or serve our students well.

In 2009, the National Governors Association (NGA) and the Council of Chief State School Officers (CCSSO) decided to address this problem. Together with 48 states, two territories, and the District of Columbia, they began the development of the CCSS in English language arts and mathematics to define consistent and clear expectations of the skills students need in order to compete in college and the workplace. The final standards were informed by nearly 10,000 public comments and by standards in other top-performing countries. These standards, released in June 2010, have already been adopted by 43 states and the District of Columbia at press time.

The assessment Consortia described in this publication will be building comprehensive assessment systems for the primary purpose of measuring student progress against these new CCSS.

How are these standards different from those most states currently have, and how will they impact instruction and the assessments used for accountability? The Center asked two thoughtful and deeply knowledgeable educators to provide their responses to these questions.

THE COMMON CORE STATE STANDARDS FOR MATHEMATICS

By William G. McCallum

In crafting a restaurant menu, locally grown produce served in regional recipes makes sense. Academic standards, however, need to travel across state lines in order to prepare students from all parts of the country to compete on the national and world stage.

In recognition of this, CCSSO and the NGA came together to work on what has come to be known as the CCSS. These standards represent the hard work and best thinking of far more people and organizations than can be listed here.[1] As a lead writer, I was fortunate to work with Jason Zimba, Phil Daro, and a distinguished group of mathematicians and math educators.

The standards began with the idea that all students will be ready for success in nonremedial college mathematics and set about teasing out what, exactly, that means for each grade level. We considered available research (scant in some areas, stronger in others); expectations about what students learn and when they learn it in other countries; and expectations of the most rigorous state standards. We avoided dictating matters such as how technology should be used or what pedagogical approach was right. We kept the focus on what, not how, students learn.

The effort paid off. The standards have received very high marks from organizations like the Fordham Foundation that rate academic standards. Students meeting these standards will be poised for success after high school.

The Standards Are Focused

The CCSS bring a new focus and coherence to the mathematics curriculum. These standards avoid the "mile wide and an inch deep" problem that has characterized American education.

In the elementary grades, more than half the time in each grade is focused on addition, subtraction, multiplication, and division (number and operations), with most of the remaining time devoted to geometry. Setting aside stale debates, the standards build both skills and understanding in parallel, each reinforcing the other, so that students retain knowledge rather than forgetting it by the next grade. For example, students are expected to know addition facts and multiplication tables from memory, even as they build an understanding of the relation between addition and subtraction and then between multiplication and division.

The Standards Are Clear and Coherent

Embedded in the standards is information that guides teachers on how mathematical knowledge builds coherently from one grade to the next. For example, the teaching of algebra in grade 8 actually starts in grade 1. This coherent stairway begins when students are asked to think algebraically about addition, subtraction, multiplication, and division. It widens in grade 3 to include fractions and decimals, reaching, in grade 6, a solid platform of understanding on which to scaffold work with expressions and equations, culminating in the study of functions in grade 8.

Providing this level of clarity helps teachers avoid re-teaching content from previous years. Because sufficient time is allocated and important ideas are developed over many years, there will be less need for teachers to repeat the same content year after year. This avoids another commonly criticized feature of today's mathematics education system: endless cycles of repetition, particularly in middle school.

Because sufficient time is allocated and important ideas are developed over many years, there will be less need for teachers to repeat the same content year after year.

[1] Detailed information on the initiative may be found at **www.corestandards.org**.

The Standards Are Rigorous

The foundations laid in elementary school allow for rich hands-on experiences in middle school with statistics and probability, ratio and proportion, geometry, and algebra — the bedrock on which high school builds the mathematical knowledge needed for today's colleges and careers and for an informed citizenry. During the middle school grades, students informally study sampling and probability, forming the foundation for a serious study of statistics in high school. This knowledge is necessary to make sense of data, to extract useful information from it — and finally — to determine if, when, and how confidently the data can be used to inform decisions in science, commerce, and society.

The high school standards define a rigorous level of college and career readiness. The melding of skills and understanding begun in the elementary grades prepares students to apply mathematics to novel situations, a key requirement for work in colleges and careers. In addition to standards for quantitative reasoning, algebra, geometry, statistics, and probability, the high school standards include an overarching set of modeling standards embedded in every other area.[2]

The Standards Include Standards for Mathematical Practice

In addition to content standards, the CCSS lay out eight standards for Mathematical Practice, which describe the ways of thinking and habits of mind of a proficient user of mathematics.

For example, Practice Standard 5, "Use appropriate tools strategically," requires that students make sound decisions about when to use technological tools and how to judge the answers they generate. Practice Standard 2, "Reason abstractly and quantitatively," resolves another sterile debate in mathematics education — an argument over the importance of pure vs. applied — by requiring both the ability to use symbols proficiently and the ability to attend to the contextual meaning of the symbols used.

The Standards Provide Priorities for Assessment

The introductions to grades K–8 identify two to four critical areas for each grade level, setting priorities for teachers, professional developers, and assessment writers. For example, in grade 3, the areas are multiplication and division, fractions, area, and two-dimensional shapes.

Faithful assessments will focus most of their time on these critical areas. They also will address the structure of mathematical knowledge by including multistep problems that require students to put together different but connected skills and understandings. They will include word problems and modeling problems that require students to read about a situation, represent it mathematically, carry out procedures for solution, and interpret the solution in terms of the context. They will attend to the standards for mathematical practice by designing tasks that draw on the habits of mind and ways of thinking of a mathematically expert practitioner. All of this will require a higher proportion, than we see today, of free-response items, assessments embedded in classroom instruction, and assessments that ask students to make a strategic choice of mathematical and technological tools.

The CCSS are only a beginning. Focused and coherent standards for mathematics can be a gift to teachers; the knowledge and skills contained therein should be thought of as our promise to students. By taking seriously the implications for curriculum and assessment, we can help teachers unwrap the gift and deliver the promise.

[2] Modeling is the process of choosing and using mathematics and statistics to represent and analyze situations in order to understand them better and to make better decisions.

William G. McCallum is a Distinguished Professor of Mathematics and head of the Department of Mathematics at the University of Arizona. He co-led the writing of the CCSS in Mathematics with Jason Zimba, professor of physics and mathematics at Bennington College, and Phil Daro, Senior Fellow, Mathematics, America's Choice, and San Francisco Director, Strategic Education Research Partnership.

THE COMMON CORE STATE STANDARDS FOR ENGLISH LANGUAGE ARTS & LITERACY IN HISTORY/SOCIAL STUDIES, SCIENCE, & TECHNICAL SUBJECTS

By Elise M. Frangos

Oliver Wendell Holmes explained, "A mind that is stretched by new experience can never go back to its old dimensions."

As a longtime English language arts teacher and Curriculum Director who has taught across grades 5–12, I think the new CCSS will shift literacy instruction toward empowering students with more of the skills they need to succeed in college and at work.

Before this happens, the CCSS will foster many spirited conversations among colleagues, between schools, and within districts. It will profoundly influence professional development and the inception of new pedagogical techniques. But ultimately, students across all economic backgrounds will be stretched in powerful and multidimensional ways.

Student literacy experiences will directly engage students in ways that are critical to their future success. The CCSS document distills support for what we already know works — that students need to interact with great texts, drill into them, comprehend and evaluate them, and use them as models for their own creative work.

Its emphasis directs students toward becoming *rhetors*, people who can speak and write effectively to communicate with others while appreciating context, understanding their audience, and knowing their purpose. In essence, the CCSS will help students find their voices and more effectively partake in both face-to-face and virtual communities.

Revitalizing Rhetoric, Promoting Fresh Writing

Whether one teaches composition in grade 2, grade 7, grade 10, or college, a common teacher lament is that student essays are often in search of a thesis. Meandering essays "talk about stuff," but students have difficulty forming an argument. The student may have written a lot, but the teacher wonders, where was she going? What was the writer's purpose? Students need to not only think and feel, they also need to question, gather evidence, shape, re-shape, and revise their understandings. They need the opportunity to formulate arguments and argue a lot. They need to know how to create claims and launch them after gathering sound, informed evidence. Students need the tools and models of civil discourse and to study examples of successful written arguments and those that failed. Through the CCSS, they will get experience in this.

The CCSS outline what students will know and be able to do. By grade 6, students will trace and evaluate an argument and determine the specific claims supported by evidence and those that are not. By grades 9 and 10, students will have read U.S. documents of historical and literary significance and will learn to delineate and evaluate the reasoning in these texts.

The study of argument won't be limited to expository writing; student writers will think expansively. Whether a student argues that Conrad's depictions of the River Thames and the Congo in *Heart of Darkness* stand in contrast, creating images of the known vs. the unknown, or if she claims that the calm, lovely natural world of Golding's island imagery in *Lord of the Flies* intensifies the horror of Piggy's death, she is still forming an argument. Effective composition, whether focused on imaginative literature-prose, poetry, or drama, is based on knowing one's purpose, content, and audience and relaying an argument in such a way that the piece hits the target.

With the advent of the CCSS, students will be taught to support claims with varied evidence, guide readers to their conclusion, and also anticipate the perspectives of those who differ with their arguments. Once students are taught the tools of rhetoric, they will see that an argument is the backbone of all expository and literary work.

The CCSS provide the backwards planning to help students get there. In the elementary years, students will focus on the comprehension of main ideas in their reading. As students progress through the curriculum, in grade 6 they will "trace and evaluate an argument and the specific claims in a text distinguishing claims that are not supported by evidence." To accomplish

this, key Aristotelian claims of ethos, logos, and pathos will have to be taught. By grade 9, students will be "able to read and comprehend seminal U.S. documents of historical and literary significance." Students also will be expected to know how to make a counterclaim or concession.

Discarding the Five-Paragraph Straitjacket

When student writers display the backbone of a solid argument, it is often supported in the form of a five-paragraph essay. This formulaic template offers the younger student a predictable skeleton for writing, but it can wrench the purpose of the composition, confine proofs to the prescribed three, and fail to engage the reader. Writers learn best from reading. The shift toward reading great nonfiction, in addition to imaginative texts, will assist student writing.

The Study of Sentence Scrambling

The CCSS extend grammar study into the realm of syntax. We may see fewer middle school, high school, and college level writers clinging to the standard sentence form of subject/predicate. By grade 5, students will "expand, combine, and reduce sentences for meaning, reader/listener interest, and style." By high school, students will " … Use words, phrases, and clauses as well as varied syntax." The focus on sentence acrobatics in the CCSS propels teachers to work with students on sentence structure and word arrangement. Sentences will be written to suit the desired musicality in writing or the purposes of the argument.

Ethical Information Gathering

Twenty-first century students need to know how to gather information and communicate information with people in front of them and beyond the classroom in fresh, clear ways. With the explosion of research sites, students need to know what research is evidence-based and salient to their research questions. The CCSS value ethical, multigenre research. Research will start in the early grades, focused on short projects to build knowledge. As students advance through the grades, they will gain experience gathering information from digital and print sources, learning how to synthesize multiple research sources and properly credit the sites they use.

Common Core Collegiality and Cooperation

Our increasingly diverse world is full of people with different ideas, histories, and cultures. Sadly the news and, sometimes, our schools are rife with stories of incivility, intolerance, and conflict. With the CCSS, practice in discourse is on the horizon. Small children will have the opportunity to learn concession and counterclaim. They'll build proficiency as collegial members of learning communities.

For me, the most heartening facets of the CCSS are the speaking and listening standards. As early as kindergarten, students will learn how to have collaborative conversations and learn to ask for help on grade-level appropriate topics.

The CCSS will stretch students to continue to read across genres, but learn how to use great stories encountered in canonical, contemporary, and multicultural literary texts to support their arguments. Students will learn how to formulate forceful and fortified arguments in writing or in speaking. Students will learn to research skillfully and write about what they find, weaving information together with artful sentences in organized compositions. Most importantly, the CCSS compel collaboration; students will know how to be smart, sound smart, and affirm the intelligent contributions of the people with whom they work or learn.

Our 21st-century students, heading to rigorous college work or the workplace, will benefit from the CCSS's shifts in literacy instruction, the classroom experiences that teachers will craft to transmit them, and the assessments that will inevitably measure these new directions.

A Ph.D. candidate in language arts and literacy at the University of Massachusetts, Lowell, Elise M. Frangos is the Director of English for the Massachusetts Math & Science Initiative, a nonprofit dedicated to increasing student success in Advanced Placement® courses.

Our 21st-century students, heading to rigorous college work or the workplace, will benefit from the CCSS's shifts in literacy instruction, the classroom experiences that teachers will craft to transmit them, and the assessments that will inevitably measure these new directions.

THE COMPREHENSIVE ASSESSMENT CONSORTIA:

PARTNERSHIP FOR ASSESSMENT OF READINESS FOR COLLEGE AND CAREERS (PARCC)

- MEMBERSHIP: 24 states and the District of Columbia serving more than 30 million K–12 students, approximately 62 percent of the nation's K–12 students
- GOVERNING STATES*: Arizona, Arkansas, the District of Columbia, Florida, Georgia, Illinois, Indiana, Louisiana, Maryland, Massachusetts, New York, Rhode Island, Tennessee
- PARTICIPATING STATES**: Alabama, California, Colorado, Delaware, Kentucky, Mississippi, New Jersey, North Dakota, Ohio, Oklahoma, Pennsylvania, South Carolina
- PROCUREMENT STATE***: Florida
- PROJECT MANAGEMENT PARTNER: Achieve
- HIGHER ED PARTNERSHIPS: More than 200 two- and four-year institutions, which typically receive 90 percent of all students across the PARCC Consortium states who enter college within two years of graduating from high school, will use the assessments as an indicator of readiness for credit-bearing entry-level courses.
- AWARD: $186 million

The previous information was accurate as of February 1, 2011.

The following summary of the PARCC assessment system has been approved by the Consortium.

The purpose of the PARCC system is to increase the rates at which students graduate from high school prepared for success in college and the workplace. To reach this goal, PARCC intends the assessments to help educators improve teacher, school, and system effectiveness by providing a wider variety of data that is useful for the purposes of analyzing effectiveness, calibrating interventions, holding school professionals accountable for student outcomes, supporting strategic management of human resources, and identifying mid-year professional development and support needs for educators. This, in turn, is intended to lead to higher levels of teacher and administrator effectiveness and faster rates of student and school improvement.

SYSTEM COMPONENTS

SUMMATIVE ASSESSMENTS FOR ACCOUNTABILITY

The PARCC assessment system will be composed of a series of summative assessments given across and at the end of the school year, as well as aligned formative assessment resources for classroom use.

Through-Course Assessments: These components are intended to focus instruction throughout the year on critical skills and concepts that are then assessed closer to the time of instruction, allowing for mid-year corrections. They will be given near the end of the first, second, and third quarters of the school year.[3]

These assessments will be given primarily on computers or other digital devices and primarily scored by computer, enabling rapid return of results. Multiple types of items will be used, including computer-enhanced items and performance tasks, to measure the full range of knowledge and skills called for in the CCSS. Results are expected to be reported within two weeks of assessment.

The purpose of the PARCC system is to increase the rates at which students graduate from high school prepared for success in college and the workplace.

* GOVERNING STATES cast decision-making votes on test design and policy.

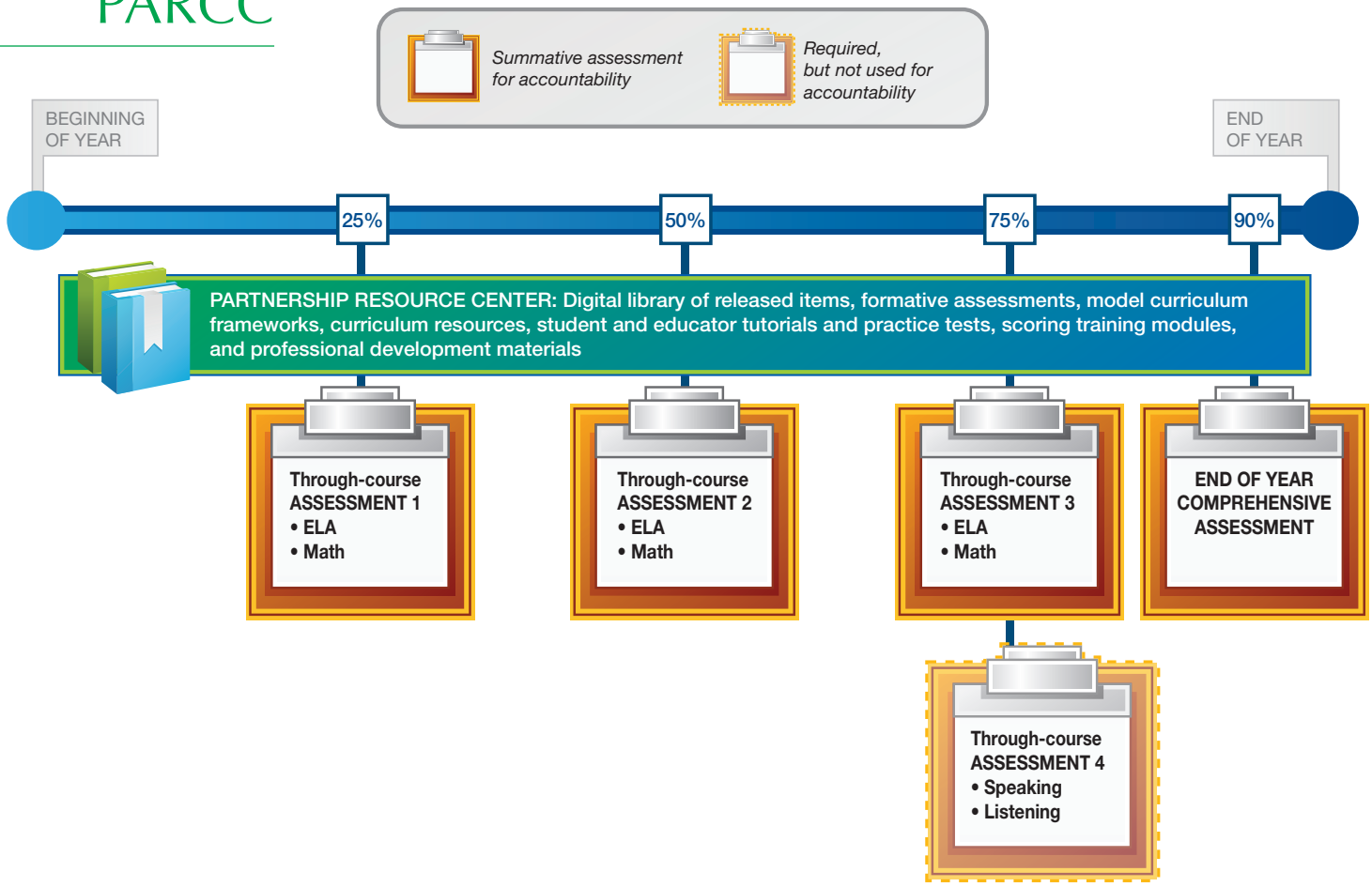
** PARTICIPATING STATES consult on test design and policy, but have no decision-making authority and must participate in pilot and field testing.

*** PROCUREMENT STATES are the fiscal agents.

[3] Specific timing of the through-course assessments will be determined after a deeper analysis of the sequencing of critical skills across the school year.

For more information about PARCC, visit **www.fldoe.org/parcc**

PARCC



English Language Arts

- **First Quarter and Second Quarter:** In a single session/class period, students will complete one or two focused literacy tasks that involve reading texts, drawing evidence from them, forming conclusions, and writing an analysis.
- **Third Quarter: Part 1:** Over several sessions/class periods, students will complete a longer written task in which they will conduct electronic searches (within a predefined set of digital sources), evaluate the quality of the sources, and compose an essay or research paper using evidence from them. At each grade level, the sources will represent a range of reading/text complexity levels to enable students at higher and lower ranges of performance to demonstrate their skills. **Part 2:** (*Required, but will not be used for accountability purposes*) Students will present their written task (above) to their classmates and respond to questions. Teachers will score the student's speaking and listening skills using a standardized rubric and may use the scores within the determination of student grades.

Mathematics[4]

- **First Quarter and Second Quarter:** In a single session/class period, students will complete one to three tasks that assess one or two of the essential topics identified in the CCSS for mathematics.
- **Third Quarter:** These items will require students to apply key mathematical concepts and processes to multistep problems. These include the ability to use technological tools to solve complex problems of the types encountered in everyday life, work, and decision making.

End-of-Year (EOY) Comprehensive Assessment: The end-of-year assessments in English/literacy and mathematics will sample all of the standards for the grade level. These assessments will be taken online during the last few weeks of the school year and will be entirely computer-scored. Each subject will be composed of 40–65 questions per subject across a range of cognitive demand.

[4] In consultation with higher education faculty, PARCC is currently looking into the best approach for the high school assessments (e.g., an end-of-course model or a cumulative end-of-domain model). It also is looking into building the high school mathematics assessments as modules, such that they can be assembled into either an end-of-course (i.e., Algebra I, Algebra II, Geometry) format or integrated course (Math I, II, III) format.

- **English/Literacy:** This assessment will focus on reading comprehension, vocabulary, and editing for grammar, usage, and language conventions.
- **Mathematics:** This component will assess conceptual understanding, procedural fluency, and problem solving. Technology will be used within items to enable students to, for example, create equations, graph functions, draw lines of symmetry, or create bar graphs.

Scoring: PARCC states will adopt a common set of performance standards and scoring rubrics so results will be comparable across states.

For the through-course assessments, a combination of computer and distributed human scoring (either teacher or vendor) will be used. To monitor the quality and reliability of scoring, 10–20 percent of randomly selected items for grade 3 through high school will be scored a second time by humans. In addition, in high school (due to the higher stakes when used to determine college course placement), an additional 10–20 percent will be scored again.

The end-of-year component will utilize 100 percent computer scoring. The Partnership plans to press for advances in automated scoring, including the use of artificial intelligence.

When paper forms are used for younger students or students with disabilities, responses will be scanned for electronic or human scoring.

The Partnership will develop a technology platform to support efficient, distributed human scoring. Member states will have the option of utilizing trained teachers (who will not score their own students' work), vendor services, or a combination thereof. In all states, all teachers will have access to the online training modules for scoring so they can more deeply understand the assessments and score classroom assignments in a consistent manner.

These scoring and administration plans may change as a result of the research conducted during the development phase.

Measuring Growth: In addition to the required assessment for grades 3 through 8 and once in high school, PARCC will develop assessments for grades 9 and 10 to support measurement of annual growth. While a specific analytic approach for calculating growth has not yet been determined, the objective will be to describe each student's relative growth, expected growth given the student's prior achievement, and the extent to which that student is "on track" toward college and career readiness. The growth measure will be reported at the student, classroom, subgroup, school, district, state, and Partnership levels.

Accountability: The Partnership plans to use the results from the through-course and end-of-year components in each subject to calculate annual combined scores for each student. A number of technical and psychometric challenges will be investigated during the development phase to determine if and how the scores from these multiple components can be aggregated to yield valid, reliable, and legally defensible scores. Both proficiency and growth data will be produced by the system for use, as needed, in accountability systems.

Reporting: An online Interactive Data Tool will provide teachers, parents, and administrators with access to results after each assessment and will include various tools for displaying data, creating customized reports, and comparing the performance of similar schools. In addition, parents will be mailed printed reports after each assessment. For administrators, the system will include tools to help identify the individual professional development needs of teachers, as well as grade-level and school-level needs.

RESOURCES, TOOLS, AND CAPACITY BUILDING

The Partnership Resource Center: This web-based platform is designed to be a continually expanding collection of resources for teachers, students, administrators, and parents. The resources, some of which will be available prior to 2014–15 to allow users to gain familiarity with the PARCC system, will include the following:

- **Model Curricular Frameworks and Exemplar Lesson Plans** — PARCC will provide support to state efforts to build these resources and will provide the digital platform for sharing them across states.
- **Released Test Items and Performance Tasks** — Teachers will be able to use these within the flow of instruction to check student understanding. States may contribute existing state-owned items or tasks aligned to the CCSS. Within a few years, all performance tasks used in the through-course summative assessments will be added, along with student performance data, scoring rubrics, and sample responses for each item. The item bank will include capabilities for sharing, improving, analyzing, comparing, ranking, and accrediting items, as well as formative and interim assessments.

- **Educator Training Materials** — Designed to help teachers understand the assessment system, implement the assessments, and interpret and use the results.
- **Online Practice Tests for Educators and Students** — These will allow teachers, students, and parents to become familiar with the assessments.
- **An Item Development Portal and Tools** — Teachers can develop their own innovative, computer-scored assessment items and share them with others via the item bank.
- **Optional Performance Tasks for Grades K–2** — Teachers and schools can use these "ready-to-use" formative tasks to monitor students' performance and progress. The tasks will consist of developmentally appropriate measures such as observations, checklists, running records, and on-demand performance events and may include the use of technology innovations, such as touch screens.

The Interactive Data Tool: See *Reporting*, on page 9.

Sharing State-Developed Tools: Formative and diagnostic tools being developed by member states and districts may be added, including a diagnostic reading tool (New York City), classroom reading and math diagnostic assessments (Pennsylvania), and an adaptive assessment platform (Tennessee).

Capacity Building: To help educators use the new assessment system well, the Partnership will:

- Build a leadership cadre of content experts within each state;
- Develop training tools to help educators implement the assessment system;
- Develop a sequence of online training modules for educators to learn to score, interpret, and use the assessment results; and
- Share advice on effective ways in which educators can understand and address the curricular and instructional implications of the CCSS and the Partnership's assessments.

TIMELINE

2010–2011	Development and approval by member states of common policies and procedures
2011–2012	Item and task development, piloting of components
2011–2012	Development of professional development resources and online platform
2012–2014	Field testing
2014–2015	New summative assessments in use
Summer 2015	Setting of achievement standards

TECHNOLOGY

Technology is a critical component for all aspects of the PARCC assessment system, from test delivery, administration, scoring, and reporting to delivery of professional development and model lesson plans. The Partnership plans to require that all of the technology created with the support of federal RTTT resources be open source, and any pre-existing technology employed in the system be either open source or documented in a fully transparent way. PARCC received a supplementary $10 million award to support development of a highly robust and stable system and to accelerate advances in technology-enhanced items and scoring engines.

Technology is a critical component for all aspects of the PARCC assessment system, from test delivery, administration, scoring, and reporting to delivery of professional development and model lesson plans.

THE COMPREHENSIVE ASSESSMENT CONSORTIA:

SMARTER BALANCED ASSESSMENT CONSORTIUM (SBAC)

- MEMBERSHIP: 31 states serving more than 20 million K–12 students, representing about 45 percent of the nation's K–12 students
- GOVERNING STATES*: Connecticut, Hawaii, Idaho, Kansas, Maine, Michigan, Missouri, Montana, Nevada, New Hampshire, New Mexico, North Carolina, Oregon, Utah, Vermont, Washington, West Virginia, Wisconsin
- ADVISORY STATES**: Alabama, Colorado, Delaware, Iowa, Kentucky, New Jersey, North Dakota, Ohio, Oklahoma, Pennsylvania, South Carolina, South Dakota, Wyoming
- PROCUREMENT STATE***: Washington
- PROJECT MANAGEMENT PARTNER: WestEd
- HIGHER ED PARTNERSHIPS: More than 170 two- and four-year institutions in these 31 states have committed to participate in the Consortium, help design the new assessments, and use the assessments as an indicator of readiness for credit-bearing entry-level courses in lieu of existing placement tests. These participating institutions typically receive 74 percent of all students in SBAC Consortium states who begin college within two years of graduating from high school.
- AWARD: $176 million

The previous information was accurate as of February 1, 2011.

The following summary of the SBAC assessment system was developed by the K–12 Center and has not been formally approved by SBAC.

The design of the SMARTER Balanced Assessment Consortium is intended to strategically "balance" summative, interim, and formative assessment through an integrated system of standards, curriculum, assessment, instruction, and teacher development, while providing accurate year-to-year indicators of students' progress toward college and career readiness.

The assessments and curricular materials will rely on research-based learning progressions, which further define how students acquire the knowledge and skills called for in the standards. Summative assessments will include both extended performance tasks and a computer-adaptive end-of-year assessment to assess the full range of the CCSS. In addition, SBAC will provide a suite of optional interim and formative tools and resources. These include: computer-adaptive interim assessments using nonsecure tasks and items of types similar to those used in the summative assessments that provide teachers with instructionally useful information about each student's progress during the year; formative tools and strategies for more regular classroom use; and professional development resources in the formative assessment process and use of assessment data of all types to adjust and improve instruction.

Technology will be leveraged in this design in several ways: adaptive testing will be used to enhance the precision of scores across the full achievement spectrum; technology-enhanced test items will expand the range of skills that can be assessed; online professional development resources and research-supported instructional tools will support improved instruction and school leadership; and, through use of an interoperable electronic platform, the Consortium will support both standardized and customized reports that can be targeted to a range of audiences for tracking and analyzing progress.

A guiding principle for the SBAC Consortium is "responsible flexibility." SBAC will make it possible for states to customize system components, while also ensuring comparability of student scores across all participating states on the summative assessments.

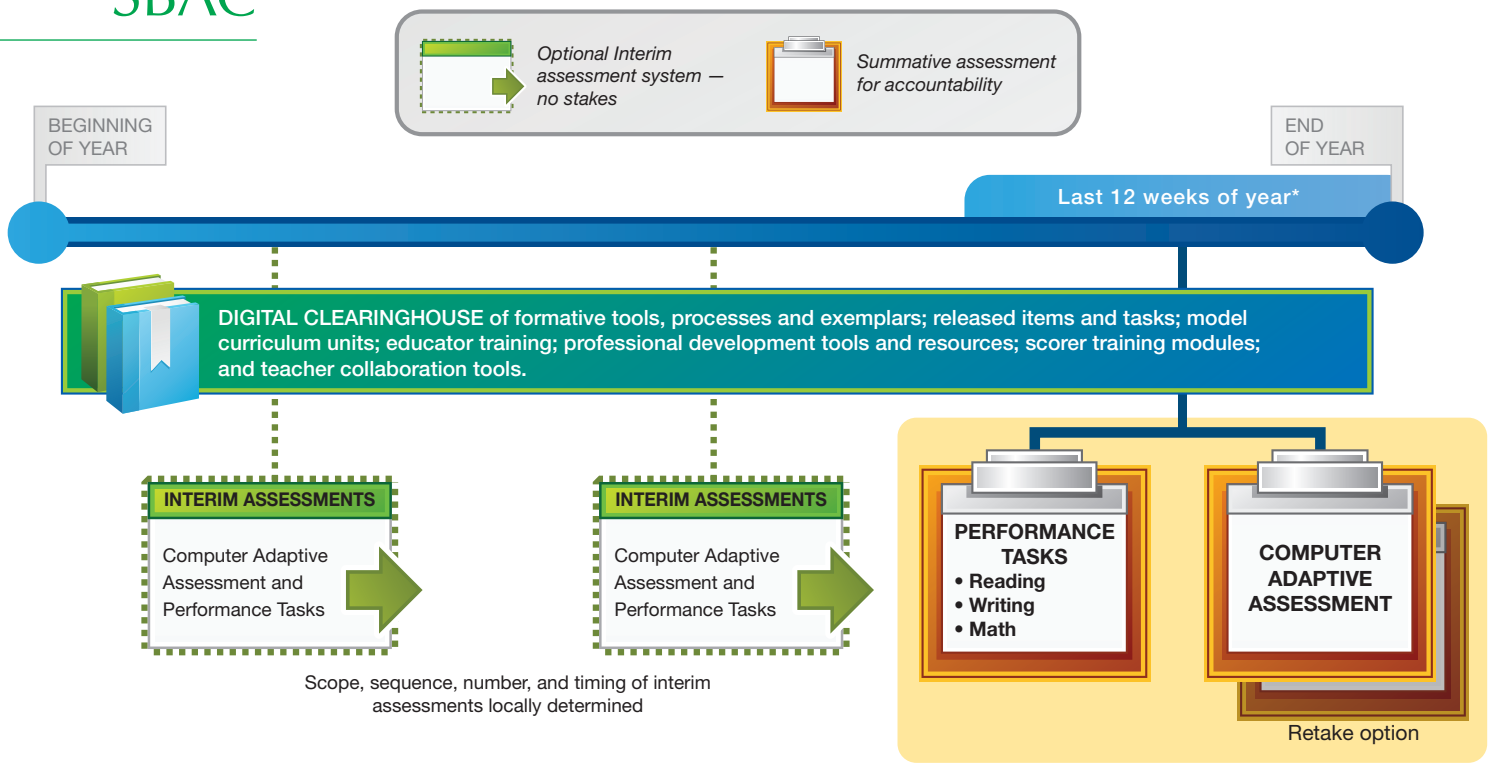
The design of the SMARTER Balanced Assessment Consortium is intended to strategically "balance" summative, interim, and formative assessment through an integrated system of standards, curriculum, assessment, instruction, and teacher development, while providing accurate year-to-year indicators of students' progress toward college and career readiness.

* GOVERNING STATES cast decision-making votes on test design and policy.

** ADVISORY STATES consult on test design and policy, but have no decision-making authority.

*** PROCUREMENT STATES are the fiscal agents.

For more information about SBAC, visit **www.k12.wa.us/smarter**



* Time windows may be adjusted based on results from the research agenda and final implementation decisions.

SYSTEM COMPONENTS

SUMMATIVE ASSESSMENTS FOR ACCOUNTABILITY

Assessments will be developed for English language arts and mathematics for grades 3–8 and 11, with assessments for grades 9 and 10 available for states that choose to use them. Although all assessments are to eventually be delivered via computer, the Consortium expects to offer a paper-and-pencil option for three years to support states that do not yet have sufficient technology infrastructure to make a complete transition at the outset.

Taken during the final 12 weeks of the school year*, the summative assessments for each grade and subject will have two major components: performance tasks and a comprehensive end-of-year computer adaptive assessment, as described below. All of these assessments will provide students with information regarding their achievement, growth, and progress toward college and career readiness by the end of high school.

- **Performance Tasks:** Students will complete one task in reading, one in writing, and two in mathematics annually[5] during a Consortium-defined testing window. Each task will be delivered via computer and will generally require one to two class periods to complete. These tasks will evaluate aspects of the CCSS that are difficult or not possible to assess through more traditional items. They will involve student-initiated planning, management of information and ideas, interaction with other materials and/or people, and production of an extended response such as an oral presentation, exhibit, product development, or an extended written piece. A combination of machine and teacher/human scoring will be used, with results available as soon as possible.

- **Computer Adaptive Assessments:** The computer adaptive component will consist of approximately 40–65 questions per content area presented within a computer-adaptive assessment. It will include selected-response, constructed-response, and technology-enhanced items. The system will use a combination of immediate scoring by computer and rapid online scoring by teachers. This component includes a retake option, as locally determined. Students who are approved to do so may take the assessment a second time, but will see a new set of items. The student's highest score would be used to determine

* Time windows may be adjusted based on results from the research agenda and final implementation decisions.

[5] This proposed number of performance tasks may be modified after further review of the CCSS.

annual achievement and annual growth relative to staying on track to college and career readiness.

The Consortium also will conduct studies to determine whether distributed summative assessments (a series of tests taken across the school year) are sufficiently valid, reliable, and comparable to the above end-of-year computer adaptive component to be offered as an alternative to it. The distributed model would likely be developed based on content clusters and would allow students to demonstrate mastery in specific units of content throughout a grade level or course.

Scoring: Performance tasks will have some components that are scored by computer and others that require human scoring. A Consortium priority is the strategic involvement of teachers in the development of items and scoring guides and in the scoring of constructed-response items (10 percent teacher scored) and performance tasks (33 percent teacher scored), although no teacher would score his/her student's responses. An online system will be developed to allow efficient distributed human scoring and monitoring of the accuracy of each reader.

For the computer adaptive component, selected-response and technology-enhanced items will be computer-scored and preliminary results from this set of items will be reported immediately. Additional items that can be reliably scored using artificial intelligence engines will be electronically scored, with 10 percent back-read by humans to verify the accuracy of the engine. Final combined scores for these summative assessments are expected to be delivered within two weeks.* The Consortium plans to leverage advances in both electronic item types and electronic scoring to support its design and will invest in the development of a training system for human scorers.

Measuring Growth: The Consortium intends to build vertical scales across the grade 3–11 span in English language arts and mathematics, which can then be used as the basis for growth measures evaluating the individual's progress toward college and career readiness across the years, but also will build horizontal (grade specific) scales. Both the summative assessment results and the interim assessment results will be reportable on these vertical and horizontal scales. The Consortium proposes to allow states flexibility in selecting the specific growth model that is appropriate for their state. The Consortium will conduct studies of the characteristics of different models when used in conjunction with the data from the summative assessments to inform subsequent decisions.

Accountability: Student scores from both the performance tasks (one in reading, one in writing, and two in math per year) and the computer adaptive assessment will be combined for the annual summative score. Research will be conducted to inform decisions concerning the aggregation and weighting of the results from these two components.

Reporting: A common electronic platform will be developed to manage assessment data and provide sophisticated data reporting and analysis tools for customized reports. Students, teachers, parents, and administrators will be given security settings to access appropriate data only. Student scores on the performance tasks will be reported separately, as well as in combination with the computer adaptive assessment. Student performance levels will be explained with examples to aid interpretation. Reports will provide item-level information for clusters of items, provided that this is found to yield valid and reliable information. In addition to summative results, scores from the interim assessments throughout the school year will be available in the same reporting suite and report more detailed information concerning progress toward that grade level's standards. This system also will include links to model curriculum and instruction resources and assessment professional development resources. The reporting tool will be customizable, allowing each state to "brand" the reporting in a manner consistent with other state-level reports.

RESOURCES, TOOLS, AND CAPACITY BUILDING

Optional Interim Assessments: These optional computer adaptive assessments can be self-administered several times each year (to be determined by states/locals) and will provide near-immediate results on the same scale as the computer adaptive component of the summative assessment. The item types will mirror those on the summative comprehensive assessment.

Two modes of test administration will be available, both of which can be given multiple times per year at the discretion of the state, district, or school. One version mirrors the length and scope of the end-of-year computer adaptive assessment and yields a scale score that can be used as a growth or achievement metric. A shorter "cluster assessment" version also will be available that assesses, at a deeper level, a smaller set of standards based on defined learning progressions, thereby providing more detailed diagnostic feedback. The items will be stored in a nonsecure item bank and can be grouped into customized clusters based on state or local curricula and can be administered before, during, or near the

* Time windows may be adjusted based on results from the research agenda and final implementation decisions.

end of instruction. Reports of student results will link teachers to appropriate formative tools and strategies for their students and professional development resources.

Comprehensive Electronic Platform: The SMARTER Balanced Assessment System will be built around a comprehensive electronic platform that contains an expanding collection of resources for teachers, administrators, students, and parents, including:

- **A System Portal** — This portal will serve as the single point of entry for educators, students, parents, and policymakers to all components of the system. In addition to the features described below, the portal will provide access to the assessment delivery platform, the distributed hand-scoring platform, virtual classrooms, and issue-focused chat rooms.
- **The Educator Dashboard** — A secure online portal will allow educators to:
  - download, view, and analyze assessment reports, scoring rubrics, and longitudinal data;
  - generate custom reports;
  - access model curricula that are aligned to the CCSS;
  - access research-based instructional strategies and interventions related to specific assessment results for individuals or subgroups;
  - access vetted instructional units, formative tools, and sample performance tasks; and
  - network with teachers to share information and resources and discuss curriculum, instruction, and assessment.
- **Formative Tools, Processes, and Practices Clearinghouse** — To be developed for grades 3–8 and high school, this bank of resources will include:
  - formative assessment tools and strategies, including the use of performance tasks to solicit formative information, and rubrics that can be used by teachers on-demand to support teaching and learning;
  - assessments created by teachers using these tools and instruments, which can be administered as computer-adaptive assessments, teacher-administered performance tasks, or classroom exercises; and
  - research-based instructional tools and processes.
- **Item Development/Scoring Application** — Online training modules will be available for both development of assessment items and tasks and for scoring of items and tasks. For those educators who successfully complete the training, item authoring and scoring software will become accessible.
- **Reporting Suite** — See “Reporting” on page 13.
- **Feedback/Evaluation Tools** — These tools will support regular surveying of system users and vetting of submitted materials.

TIMELINE

2011	Development of formative tools, processes, and practices underway Specifications for summative and interim assessments developed
2012	Summative and interim item development completed Interim item pool becomes available for use
2013	Field testing of items for adaptive summative assessment completed
2014	Preliminary achievement standards proposed and other policy definitions adopted
2015	Operational summative assessment administered Final achievement standards verified and adopted

TECHNOLOGY

The Consortium plans to develop an open-source technology platform for this assessment and learning system using a combination of existing and newly developed open-source software and proprietary software. Upon completion of the system development, a public license defining this as free, open-source software will be created.

FINDING SOLUTIONS, MOVING FORWARD

By Nancy Doorey

The assessment system designs developed by the two Consortia have some attributes in common, but also differ sufficiently to allow for "dueling banjos" during the design, development, and piloting phases — a period during which we can learn a great deal that will inform these two systems and the future of the field of assessment more broadly.

PARCC places emphasis on the use of focused, computer-delivered tasks (also called through-course assessments) that assess key skills and concepts at key points across the school year, closer to the time of instruction. SBAC places a priority on the use of adaptive interim and summative assessments to gain greater score precision and student engagement. Both will stimulate the development of a broader range of computer-delivered and -scored items and more informative reports.

Embedded in these designs, however, are a number of important and unresolved psychometric challenges. How far can we push the frontiers of measurement during this development phase? Can we find better solutions to address two priorities that stand in tension: first, the need for highly precise and reliable data for high-stakes decisions; and second, the need for assessments that require students to apply knowledge and skills to solve complex, real-world problems? Moreover, can we use this federally funded initiative to develop an information infrastructure that will allow states, districts, and schools to significantly accelerate the rate at which we can access and analyze data to propel continuous system improvement?

Over the next several years, the Center for K–12 Assessment & Performance Management at ETS (the Center) will bring together some of the best minds in the country to debate and identify promising approaches to the challenges regarding assessment and performance management systems and will share the products from these activities publicly in order to stimulate and contribute to broad advances in the field. We will begin in 2011 with two challenges that are central to both of the Consortia designs: 1) the development and use of through-course summative assessments and 2) the use of technology-enhanced tasks and scoring engines.

Through-Course Summative Assessments:

A through-course design is "one in which an assessment system component or set of components is administered periodically during the academic year" (Race to the Top Notice of Invitation to Apply, USED, 2010). This format allows for the assessment of skills and concepts that were recently taught, enabling mid-course corrections, and may include items designed to inform the next unit of instruction. The PARCC Consortium has chosen to include three through-course assessments in mathematics and in English language arts each school year. These will likely be given near the end of the first, second, and third quarters, will consist of a small number of tasks, and will be augmented by a comprehensive end-of-year assessment. The SBAC plans to develop the model shown on page 12 of this guide, which does not contain through-course assessments. However, they plan to investigate an alternative model in which the end-of-year assessment is replaced by a series of computer-adaptive assessments, each of which assesses a cluster of standards and are aggregated with the annual performance tasks.

This through-course assessment format raises a number of measurement challenges. First and foremost, the determination of the subset of skills and concepts to be emphasized in these through-course assessments requires that we identify within each content area and grade level the "keystone" topics or cognitive targets for which deep mastery is necessary for — and highly predictive of — readiness for college and a career. What are those skills and concepts, at what level of mastery, and in what sequence, if one such sequence exists? Studies will need to be carried out to gain deeper understanding than we currently have to support these decisions.

We will begin in 2011 with two challenges that are central to the Consortia designs: 1) the development and use of through-course summative assessments and 2) the use of technology-enhanced tasks and scoring engines.

Designing these components such that they can be placed onto a common scale and equated from year to year may require new approaches. Policy decisions concerning the weighting of the individual components into a composite annual score will need to be informed by data from field tests to ensure that the final composite scores are legally defensible for use in high-stakes decisions concerning individuals. Designing writing tasks that accurately reflect the skills needed in college and a career — such as locating and evaluating sources and writing a position paper supported by evidence — will require that we develop sound methodologies to extract sub-scores for reading and writing and that the content of the topic be addressed in English literacy, science or history as well.

The goal of assessing clusters of highly important competencies throughout the year is an important and worthwhile one. Finding solutions to the measurement challenges would help our schools identify individual student needs for intervention or acceleration throughout the year. The Race to the Top Assessment Program, therefore, will likely stimulate important advances in the measurement field.

Technology-Enhanced Tasks and Scoring Engines:

The demand for the transition from paper-and-pencil to online testing has grown as parents and teachers demand faster return of results and states seek to reduce the costs — in dollars and personnel time — associated with printing, shipping, and securely handling millions of printed test booklets.

Computer-based assessment also can bring benefits to the quality of the assessment itself. It creates the opportunity for an expanded set of built-in accommodations for students with disabilities and English-language learners, such as text-to-speech utilities, text enlargement, and language translations. Many skills and competencies that are fundamental to college and career readiness, and are called for within the CCSS, cannot be assessed on paper, such as online searching (in contained digital libraries) and the use of word processing and data analysis software. New interactive computer tasks allow us to gain information about both the content knowledge of the student and the processes used by the student in solving complex problems. Here, the more challenging electronic and online games can provide some insight into what may be possible to incorporate into simulations that require the application of knowledge and skills to real-world problems.

There are clearly new frontiers to be reached in the area of automated scoring as well. A stated goal of the PARCC Consortium is to drive innovation in this area such that all items and tasks can be scored very quickly by computer. Artificial intelligence engines exist that score the large majority of student essays, at least as reliably as humans, and "send back" those essays that are so unique or creative as to require human scoring. However, as we look to assess writing in the context of science, English literature, or history, as called for in the CCSS, new advances are needed to produce reliable sub-scores for both writing and the content area constructs assessed.

The development of interactive computer tasks and their automated scoring engines is challenging, particularly when used within high-stakes assessments. To illustrate, the National Board of Medical Examiners has been working on interactive medical case simulations and automated scoring methodologies for use in their licensure examination for nearly 30 years and using them in operational exams for more than a decade. Examinees currently complete nine simulations, each of which takes about 25 minutes. However, to ensure high comparability from one administration of the test to the next, each examinee also takes approximately 480 selected response items. If results from the new K–12 assessment systems are to be used for comparable high-stakes decisions regarding individuals, such as the awarding of high school diplomas, similarly high thresholds for psychometric quality must be met. This is clearly a time for aggressive research and development if we are to realize and use the benefits of such new assessment items.

We must utilize this opportunity to create the infrastructure that will enable us to maximize the potential of each child, accelerate our learning, and regain our former position as first in the world in educational attainment and equity.

Looking Beyond 2015:

Some of the challenges discussed previously may be solved within the next two or three years, in time for application within the initial roll-out of PARCC and/or SBAC. Others will require more time to develop prototypes, pilot, and prepare for use within high-stakes assessments. New technologies will be developed, which will impact the technical and financial feasibility of some approaches, and open the doors to new approaches.

This points out the need to think of the next four years as only the beginning — the development of a strong foundation — for a new, robust data and research platform from which we can leverage technologies to accelerate advances in K–12 education and to enhance student learning for all children. The CCSS and aligned common assessments create the opportunity to shift from norm-based educational decision making — identification of what works for most students, most of the time — to more nuanced and personalized educational decision making.

The data platforms of 2015 should allow us to answer questions such as: What would work best for this student based on accumulated information from large numbers of students with similar prior achievement patterns? What are the most effective professional development activities for the unique needs of this particular teacher as identified by the progress of his/her students?

To realize this vision, we will need to standardize data formats such that anonymized student, teacher, and assessment data can be collected in real time and organized to support timely analysis and research, and that high-quality digital curriculum and assessment content can be similarly accessed and utilized by all.

We are truly at a unique moment in the history of American public education. We must utilize this opportunity to create the infrastructure that will enable us to maximize the potential of each child, accelerate our learning, and regain our former position as first in the world in educational attainment and equity.

Nancy Doorey is the Director of Programs at the Center for K–12 Assessment & Performance Management at ETS. She is a doctoral candidate (Ed.D., ABD) in educational leadership at Teachers College, Columbia University.

Let's Animate the Discussion!

The Center has developed an animated slide presentation that walks viewers through each element of the Consortia designs for their assessment systems and related support materials. This presentation has been approved by each Consortium. We hope you will find it useful.

Visit **www.k12center.org/publications.html** to download it, and be sure to view it in slideshow mode.

You can sign up to receive future updates and materials from the Center by sending an e-mail to **mail@k12center.org**.

Center for K–12 Assessment & Performance Management at ETS

A catalyst and resource for the improvement of measurement and data systems to enhance student achievement.

The Center will work with nationally recognized measurement experts from across the country to explore possible solutions to the measurement challenges inherent in the designs of the new assessments and will share the resulting ideas and recommendations through webinars and our website. To sign up for notices as these resources are made available, visit

www.k12center.org

Created by Educational Testing Service (ETS) to forward a larger social mission, the Center for K–12 Assessment & Performance Management at ETS has been given the directive to serve as a catalyst and resource for the improvement of measurement and data systems to enhance student achievement.